Lab 3 Practice: Summarizing and Visualizing Data

Background

Alzheimer's Disease (AD) is a serious mental illness that affects an estimated 5.3 million Americans; it is the most common cause of dementia among the elderly. Characterized by a progressive cognitive decline, AD has been notoriously difficult to diagnose due to symptom-overlap with other mental disorders; until recently, AD could only be confirmed posthumously. Researchers are investigating identifying biomarkers in interdisciplinary research involving biostatisticians, neuroscientists, geneticists, and clinicians, among other experts.

The Data

The Alzheimer's Disease Neuroimaging Initiative (ADNI) is a longitudinal study that began in 2005, and is designed to track AD biomarkers, identify at-risk patients, and evaluate the efficacy of novel treatments. The study consists of healthy individiduals (the control group) as well as adults with early Alzheimer's Disease (AD). More details can be found on the website *http://adni.loni.usc.edu/about/*. The data set provided is `ADNI`, which contains information on 276 subjects and 7 variables. The mini mental state exam is a 30 question assessment commonly used to assess cognitive impairment. The Alzheimer's Disease Assessment Scale is a more comprehensive measure of cognitive impairment. The apolipoprotein E (APOE) gene, on chromosome 19, has variants associated with high risk of AD. In this lab, your goal is to provide summary statistics and visualizations of the data.

`DX`	Alzheimer's disease diagnosis
	    AD - Alzheimer's disease
	    MCI - Mild cognitive impairment
	    Normal - Normal cognitive function

`AGE`	Age (years)

`APOE4`	Type of APOE4 variant (genetics)
	    0 - No copies of the ApoE4 allele
	    1 - One copy of the ApoE4 allele
	    2 - Two copies of the ApoE4 allele

`GENDER`	Patient gender
`MMSE`	Mini Mental State Exam (score out of 30, lower scores indicate more cognitive impairment)
`adas`	Alzheimer's Disease Assessment Scale (larger scores indicate greater dysfuction)
`WholeBrain`	Brain volume (mm^3)

Practice

1. Open the `R Reference Guide` from the `Lab Content` area of PolyLearn.

2. Open `RStudio` and then open a brand new `Rmarkdown` document by clicking on the green plus sign on the top left of RStudio. Delete everything in the Rmarkdown document. Identify the `Rmarkdown` tab of your `R Reference Guide`. Copy and paste the three code chunks (header, set up chunk, and import data code chunk) into your Rmarkdown document.

3. Identify the `Lab Data Sets` folder on PolyLearn and download the `adni` data set to a location on your computer (i.e., desktop, STAT 217 folder). Follow the steps in the `Importing` tab of the `R Reference Guide` to import the `adni` data set and save your import code in an R chunk.

4. Use an R command to determine how many observations and variables are in the data set.

5. For each variable in the data set, indicate if the variable represents a categorical or quantitative variable in reality (no R command needed, this is just your interpretation of the data).

6. Produce a `summary` of the data set. Describe why the `APOE4` variable is not summarized correctly.

7. Submit the following code to recode the `APOE4` variable to a factor, and then produce another summary of the data set. Which APOE4 genetic variant is the least common?

```
ADNI$APOE4 <- factor(ADNI$APOE4,labels=c("no copies","one copy","two copies"))
```

8. What is the median value of Mini Mental State Exam (`MMSE`)? Provide an interpretation of this value.

9. What is the mean and the median of the Alzheimer's Disease Assessment Scale (`adas`)? What does the relationship between the mean and the median indicate about the direction of the skew between the variables, and why? Produce a histogram of `adas` to verify your answer.

10. Produce an appropriate figure to visualize the relationship between Alzheimer's diagnosis (`DX`) and the cognitive impairment test `MMSE`. Describe the trend that you observe.

11. Produce summary statistics of the `MMSE` score by Alzheimer's diagnosis (`DX`). What is the average `MMSE` score for each diagnosis group? This should be reflective of the figure you produced in the previous question.

12. Produce an appropriate figure to visualize the relationship between the two cognitive impairment test scores `MMSE` and `adas`. Describe the trend that you observe.

13. Produce a contingency table showing the relationship between Alzheimer's diagnosis (`DX`) and the genetic variant `APOE4`. Among patients with a normal diagnosis, what percent have no copies of the ApoE4 allele? Among patients with a Alzheimer's disease, what percent have no copies of the ApoE4 allele? (Use R as a calculator to show your calculations in a R chunk.)

14. Submit this lab assignment as an **html** compiled from Rmarkdown. Make sure names of **all** present lab group members who contributed to this lab assignment are listed on the html file.